



UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
10/802,479	03/16/2004	Jun Wu	16113-615001	4107
26192 7590 03/06/2008 FISH & RICHARDSON P.C. PO BOX 1022 MINNEAPOLIS, MN 55440-1022			EXAMINER HAN, QI	
			ART UNIT 2626	PAPER NUMBER
			MAIL DATE 03/06/2008	DELIVERY MODE PAPER

Please find below and/or attached an Office communication concerning this application or proceeding.

The time period for reply, if any, is set in the attached communication.

Office Action Summary	Application No. 10/802,479	Applicant(s) WU ET AL.	
	Examiner Qi Han	Art Unit 2626	

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --

Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE 3 MONTH(S) OR THIRTY (30) DAYS, WHICHEVER IS LONGER, FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133). Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

Status

- 1) ☒ Responsive to communication(s) filed on 21 November 2007.
- 2a) ☐ This action is **FINAL**. 2b) ☒ This action is non-final.
- 3) ☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

Disposition of Claims

- 4) ☒ Claim(s) 1-47 is/are pending in the application.
- 4a) Of the above claim(s) 18-47 is/are withdrawn from consideration.
- 5) ☐ Claim(s) _____ is/are allowed.
- 6) ☒ Claim(s) 1-17 is/are rejected.
- 7) ☐ Claim(s) _____ is/are objected to.
- 8) ☐ Claim(s) _____ are subject to restriction and/or election requirement.

Application Papers

- 9) ☒ The specification is objected to by the Examiner.
- 10) ☐ The drawing(s) filed on _____ is/are: a) ☐ accepted or b) ☐ objected to by the Examiner.
Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).
Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).
- 11) ☐ The oath or declaration is objected to by the Examiner. Note the attached Office Action or form PTO-152.

Priority under 35 U.S.C. § 119

- 12) ☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).
- a) ☐ All b) ☐ Some * c) ☐ None of:
- ☐ Certified copies of the priority documents have been received.
 - ☐ Certified copies of the priority documents have been received in Application No. _____.
 - ☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).
- * See the attached detailed Office action for a list of the certified copies not received.

Attachment(s)

- | | |
|--|---|
| 1) <input checked="" type="checkbox"/> Notice of References Cited (PTO-892) | 4) <input type="checkbox"/> Interview Summary (PTO-413)
Paper No(s)/Mail Date. _____ |
| 2) <input type="checkbox"/> Notice of Draftsperson's Patent Drawing Review (PTO-948) | 5) <input type="checkbox"/> Notice of Informal Patent Application |
| 3) <input type="checkbox"/> Information Disclosure Statement(s) (PTO/SB/08)
Paper No(s)/Mail Date _____ | 6) <input type="checkbox"/> Other: _____ |

DETAILED ACTION

Response to Amendment

1. This communication is responsive to the applicant's response (to the restriction requirement) filed on 11/21/2007.

Election/Restrictions

2. Applicant's election without traverse of invention Group I, claims 1-17 in the reply filed on 11/21/2007 is acknowledged.
3. Claims 18-44 withdrawn from further consideration pursuant to 37 CFR 1.142(b) as being drawn to a nonelected invention Groups II and III, there being no allowable generic or linking claim. Election was made **without** traverse in the reply filed on 11/21/2007.

Specification and Drawing

4. The disclosure is objected to because of the following:
 - a. in paragraph 31, the content "...group the remaining new character strings...into 7 sets of new character strings" is unclear, because the context lacks description/definition of what the 7 sets really are and/or what the criteria/categories of the sets are used for grouping.Appropriate correction/clarification is required, without adding new matter.

Claim Rejections - 35 USC § 103

The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all obviousness rejections set forth in this Office action:

(a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102 of this title, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negated by the manner in which the invention was made.

5. Claims 1-3, 5, 9-12 and 14 are rejected under 35 U.S.C. 103(a) as being unpatentable over BADINO (US 2007/0118346) in view of LEE et al. (US 7,165,019) hereinafter referenced as LEE.

As per **claim 1**, BADINO discloses 'automatic segmentation of texts comprising chunks without separators' (title), comprising:

"extracting unknown character strings from a set of Chinese inputs", (p(paragraph)29, 'Mandarin Chinese language'; p40, 'input (Chinese) text is subdivided into syntagms' that 'is a portion of text (unknown character strings)', 'each syntagm is sent...to the segmentation module'; p85, 'each single syntagm (unknown character string) is extracted', 'decomposition into words'; p73, 'segmenting unknown words (also read on unknown character strings)' that 'are not included in the training corpus (a set of Chinese inputs)', which implies a training phase in which the system has been trained by using the corpus as input for extracting the strings, in the same/similar manner as text phase);

"determining valid words from the unknown character strings" (p46-p71) by "comparing frequencies of occurrence of the unknown character strings with frequencies of occurrence of individual characters of the unknown character string" (p22-p23, 'maximum marching (comparing) rule' as a general rule, 'the probability (corresponding frequencies of occurrence) to

that a give sequence of ideograms (unknown character string) belongs to a single word (valid word) within the lexicon (dictionary) is higher (implying a difference) than the probability that such a sequence corresponds to a plurality of shorter words (individual characters or words) concatenated within the text'); and

“generating a transition matrix [of conditional probabilities] for predicting a word given a context” (p34, ‘all the decompositions... be mapped in a lattice or matrix where each element is comprised of a word plus the respective cost’; p44-p45, ‘a sort of lattice or matrix is created (generated)’, including ‘a unitary length word, then the word with the subsequent ideogram (for predicting a word) and so on up to a give length’ (so as being interpreted as transition matrix).

BADINO does not expressly disclose the matrix being “of conditional probabilities”. However, the feature is well known in the art as evidenced by LEE who discloses ‘language input architecture for converting one text form to another text form with modeless entry’ (title), comprising ‘language model (e.g. a Chinese language model)’ (col. 6, lines 5-7), ‘which measures the priori probability of any give string of words’, ‘building a statistical language model’ by using ‘N-gram language model (such as N-gram Markov model)’ that ‘counts the number of occurrences (frequencies) of a particular item (word, character, etc.) in a string’ and ‘to calculate the probability’, utilizing ‘a large training corpus’ and ‘pre-defined lexicon (dictionary)’, and ‘predict the probability (including conditional probabilities) of a sequence of words’ (col. 10, line 48 to col. 11, line 16, and equation 1). Therefore, it would have been obvious to one of ordinary skill in the art at the time the invention was made to modify BADINO by providing statistical language model (such as n-gram language model) with the probability (including conditional probabilities) of a sequence of words, as taught by LEE, for the purpose (motivation)

of predicting the next character and/or achieving higher accuracy for the text (LEE: col. 11, line 4 and col. 5, lines 55-56).

As per **claim 2** (depending on claim 1), the rejection is based on the same reason described for claim 1, because the rejection for claim 1 covers the limitation(s) of claim 2.

As per **claim 3** (depending on claim 1), the rejection is based on the same reason described for claim 1, because the rejection for claim 1 covers the limitation(s) of claim 3.

per **claim 5** (depending on claim 3), the rejection is based on the same reason described for claim 1, because the rejection for claim 1 covers the limitation(s) of claim 5.

As per **claim 9**, it recites a computer program product. The rejection is based on the same reason described for claim 1, because the claim recites the same or similar limitations as claim 1.

As per **claims 10-12 and 14**, they recite a system. The rejection is based on the same reason described for apparatus claims 1-3 and 5 respectively, because the claims recite the same or similar limitations as claims 1-3 and 5 respectively.

6. Claims 4, 7-8, 13 and 16-17 are rejected under 35 U.S.C. 103(a) as being unpatentable over BADINO in view of LEE applied to claim 1, 3, 10, 12, and further in view of LEE (US 2004/0215465) hereinafter referenced as LEE2.

As per **claim 4** (depending on claim 3), even though BADINO in view of LEE discloses “the n-gram counts include the counts of n-tuples of adjacent [and non-adjacent] words in the set of Chinese inputs” (BADINO:p50-p52; LEE: col. 11, lines, 3-16), BADINO in view of LEE does not expressly disclose the counts of n-tuples of “**non-adjacent**” words. However, the

feature is well known in the art as evidenced by LEE2 who discloses 'method for speech-based information retrieval in mandarin Chinese', comprising 'a whole class of syllable-level indexing terms' including 'overlapping syllable segments 'with length N' (adjacent n-tuples) and 'syllable pairs separate by n syllables' (non-adjacent n-tuples) (p14 and Fig.1), which is also applied to 'character- and word-level information' for 'text queries' and 'text information records' (p17 and Fig. 2). Therefore, it would have been obvious to one of ordinary skill in the art at the time the invention was made to modify BADINO in view of LEE by providing n-gram counts of both overlapping adjacent and non-adjacent words, as taught by LEE2, for the purpose (motivation) of improving segmentation approaches and/or obtaining better retrieval results for the text (LEE2: abstract and p6).

As per **claim 7** (depending on claim 1), the rejection is based on the same reason described for claim 4, because the rejection for claim 4 covers the limitation(s) of claim 7.

As per **claim 8** (depending on claim 7), BADINO in view of LEE and LEE2 further discloses "the set of Chinese inputs includes a set of user Chinese queries to a web search engine" (LEE: col. 6, lines 30-61, 'language input system may be practiced in distributed computing environment' through a communication network (e.g. 'LAN, internet (web), etc.)'), using 'search engine'; LEE2: p2 and p4, 'multi-media information on the Internet', 'the information records keep on growing very fast on the Internet every day'; Fig. 2, 'text queries'; it would have been obvious to one of ordinary skill in the art at the time the invention was made to recognized that text queries could use web search engine because the search engine for the queries (input) could be performed in distributed environment through internet (web)).

As per **claims 13 and 16-17** (depending on claim 10), the rejection is based on the same reason described for apparatus claims 4 and 7-8 respectively, because the claims recite the same or similar limitations as claims 4 and 7-8 respectively.

7. Claims 6 and 15 are rejected under 35 U.S.C. 103(a) as being unpatentable over BADINO in view of LEE applied to claims 1 and 10, and further in view of NIE et al. (“unknown word detection and segmentation of Chinese using statistical and heuristic knowledge”, communications of COLIPS, vol. 5. NO 1&2, DEC 1995, page 47-57) hereinafter referenced as NIE.

As per **claim 6** (depending on claim 3), even though BADINO in view of LEE discloses “the frequency of occurrence of the unknown character string as compared with frequencies of occurrence of the individual characters of the unknown character string is greater” (see rejection for element 2 of claim 1), BADINO in view of LEE does not expressly disclose “wherein an unknown character string is determined to be a valid new character string” based on “a predetermined threshold”. However, the feature is well known in the art as evidenced by NIE who discloses ‘unknown word detection and segmentation of Chinese using statistical and heuristic knowledge’ (title), comprising ‘procedure for eliminating n-gram overlapping’ if ‘an n-grams contained within longer n-grams’ that ‘have a high probability of being words (with high frequency)’, and ‘an n-grams (an unknown character string) having a frequency higher than a certain (predetermined) threshold is considered (determined) as a new word (valid new character string)’ (pages 52-53, section 3.3). It is noted that one of ordinary skill in the art would have readily recognized that the above eliminating n-gram overlapping process disclosed by NIE

would satisfy and/or fall within the same scope of, the maximum matching general rule disclosed by BADINO (p22). One of ordinary skill in the art would have also recognized that the threshold in NIE would imply a frequency range that is higher than the frequency (or frequencies) of the shorter n-grams and lower than the longer n-grams in order to increase robustness; and other suitable threshold(s) could also be used for new word detection, such as using a threshold based on difference between the frequency of longer n-gram and frequency (or frequencies) of shorter n-gram(s) being within the longer n-gram (wherein the difference is implied in the teachings of both BADINO and NIE's), which achieves the same or similar predictable feature and goal (results). Therefore, it would have been obvious to one of ordinary skill in the art at the time the invention was made to combine features including the maximum matching general rule disclosed by BADINO in view of LEE, and the new word detection from n-grams using a threshold taught by NIE, to provide new word (valid new character string) detection by using a suitable threshold based on difference between the frequency of a longer n-gram and frequency (or frequencies) of shorter n-grams being within the longer n-gram, for the purpose (motivation) of better improving unknown word/character detection and segmentation of Chinese text (NIE: title and page 48, left col. paragraphs 4-5).

As per **claim 15** (depending on claim 10), the rejection is based on the same reason described for claim 6, because the rejection recites the same or similar limitation(s) as claim 6.

Conclusion

8. Please address mail to be delivered by the United States Postal Service (USPS) as follows:

Mail Stop ____

Application/Control Number:
10/802,479
Art Unit: 2626

Page 9

Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

or faxed to: 571-273-8300, (for formal communications intended for entry)

Or: 571-273-8300, (for informal or draft communications, and please label
"PROPOSED" or "DRAFT")

If no Mail Stop is indicated below, the line beginning Mail Stop should be omitted from the address.

Effective January 14, 2005, except correspondence for Maintenance Fee payments, Deposit Account Replenishments (see 1.25(c)(4)), and Licensing and Review (see 37 CFR 5.1(c) and 5.2(c)), please address correspondence to be delivered by other delivery services (Federal Express (Fed Ex), UPS, DHL, Laser, Action, Purolator, etc.) as follows:

U.S. Patent and Trademark Office
Customer Window, Mail Stop _____
Randolph Building
Alexandria, VA 22314

Any inquiry concerning this communication or earlier communications from the examiner should be directed to Qi Han whose telephone numbers is (571) 272-7604. The examiner can normally be reached on Monday through Thursday from 9:00 a.m. to 7:30 p.m. If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, Richemond Dorvil, can be reached on (571) 272-7602.

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Inquiries regarding the status of submissions relating to an application or questions on the Private PAIR system should be directed to the Electronic Business Center (EBC) at 866-217-9197 (toll-free) or 703-305-3028 between the hours of 6 a.m. and midnight Monday through Friday EST, or by e-mail at: ebc@uspto.gov. For general information about the PAIR system, see <http://pair-direct.uspto.gov>.

QH/qh
January 28, 2008



1/28/08